# Translating Player Dialogue into Meaning Representations Using LSTMs

James Ryan, Adam James Summerville, Michael Mateas,
and Noah Wardrip-Fruin

Expressive Intelligence Studio
University of California, Santa Cruz
{jor, michaelm, nwf}@soe.ucsc.edu, asummerv@ucsc.edu

**Abstract.** In this paper, we present a novel approach to natural language understanding that utilizes *context-free grammars* (CFGs) in conjunction with *sequence-to-sequence* (seq2seq) *deep learning*. Specifically, we take a CFG authored to generate dialogue for our target application, a videogame, and train a *long short-term memory* (LSTM) *recurrent neural network* (RNN) to translate the surface utterances that it produces to traces of the grammatical expansions that yielded them. Critically, we already annotated the symbols in this grammar for the semantic and pragmatic considerations that our game's dialogue manager operates over, allowing us to use the grammatical trace associated with any surface utterance to infer such information. From preliminary offline evaluation, we show that our RNN translates utterances to grammatical traces (and thereby meaning representations) with great accuracy.

## 1 Introduction

While conversational agents in service applications have become an increasingly common part of everyday life, few videogames have featured freeform conversational interaction with non-player characters (NPCs)—here, *Façade* is the only major example to date [4]. This is likely due to fundamental differences between the patterns of interaction germane to service conversational agents relative to those we envision for NPCs. In service dialogue systems, interaction is constrained and highly structured, lending well to *rule-based approaches* to natural language understanding (NLU). Contrarily, conversational interaction with ideal, futuristic NPCs would be less constrained and more open-ended, making the paradigm less suitable for rule-based approaches, since a huge number of matching rules would have to be authored to cover the larger conversational domains. *Façade*, whose NLU system *is* rule-based, partly wrangles this problem by constraining the conversational domain according to a strong dramatic progression. Still, its authors tasked themselves with producing 6,800 rules over the course of hundreds of person hours, and then relied on the additional measure of rules being promiscuous in their mapping to discourse acts [4]. As such, it is not surprising that, nearly fifteen years since its first reporting in the literature, very few practitioners of entertainment-based interactive media have taken

on the massive authorial burden requisite to employing *Façade*'s demonstrated technical approach [3]. Additionally, we note that the prospect of taking this approach would be even more daunting in interactive media lacking strong dramatic progression, *e.g.*, open-world games. Further, the rules themselves can be difficult for naive authors—*e.g.*, dialogue authors working on teams developing videogames—to compose. Finally, beyond authorial burden, there is the basic problem that matching rules, even fuzzy ones, are often brittle.

In this paper, we present a method for NLU that is intended to be less authorially intensive, less confounding to naive authors, and less brittle than rule-based approaches. This method utilizes context-free grammars (CFGs) in conjunction with the *long short-term memory* (LSTM) *recurrent neural network* (RNN) architecture. Specifically, a (potentially naive) author specifies a CFG (using a tool we have developed called Expressionist [7]) whose terminal derivations are surface utterances and whose nonterminal symbols are annotated by the author to capture semantic and pragmatic considerations [5]. Training data is then generated from this CFG, in the form of utterances paired with traces of the grammatical expansions that produced them. The learning task, then, is one of *sequence-to-sequence* (seq2seq) translation, in which we train an RNN to translate surface utterances into grammatical traces. Crucially, because the symbols in these traces have been annotated with semantic and pragmatic information, we can infer such information from any trace that the RNN translates a surface utterance into, thereby decoding the utterance into a meaning representation.

We are currently employing this method in a game that we are developing, called *Talk of the Town* [8], by having a trained RNN translate arbitrary player dialogue to grammatical traces, which are then used to procure semantic and pragmatic information that is fed to the game's dialogue manager. While we are not yet poised to explicitly compare our method to rule-based systems in terms of authorial burden, amenability to naive authors, or brittleness, we do demonstrate its accuracy in translating from surface utterances to grammatical traces (which point directly to semantic and pragmatic mark-up); additionally, we will attempt to qualitatively argue for the advantages of our approach, relative to rule-based systems, along those criteria. For more information about this project, please see our longer technical report on the subject [9].

## 2 Method

Our training data is a CFG that had already been authored—using a tool we have developed called Expressionist [7]—for the purpose of generating NPC dialogue in *Talk of the Town* [5]. In this grammar, the mark-up attributed to nonterminal symbols corresponds to the semantic and pragmatic concerns that the game's dialogue manager operates over, described at length in [6]. Using Expressionist, our grammar took approximately twenty hours for a single author to produce; it comprises 217 nonterminal symbols and 624 production rules, and is capable of yielding a total 2.8M surface utterances. The actual training data used for training our RNN consists of pairs of surface utterances matched with traces

of the grammatical expansions that produced them. To produce this training data, we sampled 5,000 surface utterances for each unique possible meaning (as determined by annotations on the symbols expanded to produce them), producing a total 345,000 utterance–trace pairs. Additionally, we utilized a denoising component that augmented these pairs with new pairs whose utterances were automatically corrupted.

For our seq2seq learning procedure, we used the Tensorflow framework [1] to develop a mapping from surface utterances to traces of the grammatical expansions that produced them. This procedure can be thought of as a translation task: the neural network translates from one language (surface utterances) to another language (grammatical traces), where instances of each language are essentially just strings. For instance, the string `Oh, greetings, Andrew.` in the utterance language translates to the string `greet(greet back(use interlocutor first name))` in the trace language. Our network utilizes LSTM cells [2], a modification of the standard RNN approach that represents the current state of the art for sequence processing. For a more detailed explanation of this aspect of the project, see our longer technical report [9].

After training the RNN, we incorporated it into the software framework that underpins *Talk of the Town*. As described in [6], conversation in our game is turn-based, with turns being allocated by the dialogue manager. When a turn has been given to a player character, the player is asked to furnish her character's next utterance. Once the player has submitted this, the dialogue manager passes the utterance to the RNN, which tokenizes it and performs seq2seq translation on it to produce a grammatical trace composed of symbols in our Expressionist grammar. From here, the dialogue manager collects all the mark-up associated with all the symbols appearing in the trace, and treats this as the meaning of the player utterance (which it processes to update the conversation state).

## 3 Preliminary Evaluation

We carried out a preliminary offline evaluation procedure that demonstrates the accuracy of our system in mapping from surface utterances to grammatical traces. To conduct this experiment, we randomized our set of training data, split it into eleven pieces, and for each piece, performed 10-fold cross validation on the remainder of the set before finally using the held-out piece as a test set. We then calculated perplexity values: both cross-validation perplexity and test perplexity averaged 1.046 across all folds, with no fold showing perplexity worse than 1.053. Low perplexity values near 1 showcase the ability of the system to translate from surface utterances to grammatical traces (and thereby semantic and pragmatic information, as explained above) nearly perfectly in this task. Further, these preliminary results indicate that this approach is robust to variations and gaps in the data, with no fold performing drastically better or worse than any other. In addition to this experiment, in [9] we provide informal results in the form transcriptions of sample conversations between us and NPCs in our game, which were made possible by the technique we describe here.

## 4 Discussion and Future Work

While we have demonstrated the accuracy of this system in mapping surface utterances to grammatical traces (and thereby semantic and pragmatic information characterizing the utterances), we would like to informally discuss the advantages of our method relative to rule-based approaches to NLU. First, we believe that our approach incurs less authorial burden, simply by virtue of the combinatorial explosion that characterizes generative grammars. This is demonstrated in the large number of terminal derivations that our grammar can generate. Further, we contend that our approach is more amenable to naive authors who might like to feature NLU in their applications. Rather than authoring procedural rules, by our approach an author uses the Expressionist graphical user interface, which is designed for naive authors and supports live feedback showing surface utterances and their corresponding annotations [7]. While training a neural network is certainly not practical for naive authors, we plan to eventually support black-box RNN training as a service associated with Expressionist. Finally, we posit that intuitively our model should be less brittle than rule-based systems. While rules in such systems work by matching discrete authored patterns (which of course may be fuzzy) against user utterances, a neural network does something similar, but with patterns at arbitrary granularities, with hierarchies (patterns of patterns) that are learned dynamically. Of course, one tradeoff here is that human-authored rulesets are much more interpretable than RNNs.

While our preliminary evaluation is promising, we are currently planning a study with actual players so that we may better understand both the successes and limitations of our neural approach. Finally, we again invite the reader to see our longer technical report on this stage of our project [9].

## References

1. Abadi, M., et al.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), http://tensorflow.org/, software available from tensorflow.org
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation (1997)
3. Lessard, J.: Designing natural-language game conversations. In: Proc. DiGRA-FDG (2016)
4. Mateas, M., Stern, A.: Natural language understanding in Façade: Surface-text processing. In: Proc. TIDSE (2004)
5. Ryan, J., Mateas, M., Wardrip-Fruin, N.: Characters who speak their minds: Dialogue generation in Talk of the Town. In: Proc. AIIDE (2016)
6. Ryan, J., Mateas, M., Wardrip-Fruin, N.: A lightweight videogame dialogue manager. In: Proc. DiGRA–FDG (2016)
7. Ryan, J., Seither, E., Mateas, M., Wardrip-Fruin, N.: Expressionist: An authoring tool for in-game text generation. In: Interactive Storytelling (2016)
8. Ryan, J.O., Summerville, A., Mateas, M., Wardrip-Fruin, N.: Toward characters who observe, tell, misremember, and lie. In: Proc. Experimental AI in Games (2015)
9. Summerville, A.J., Ryan, J., Mateas, M., Wardrip-Fruin, N.: CFGs-2-NLU: Sequence-to-sequence learning for mapping utterances to semantics and pragmatics. University of California, Santa Cruz, Tech. Rep. UCSC-SOE-16-11 (2016)